

Машинный перевод без параллельного текста

Мазуров Михаил

October 2018

1 Abstract

Задача машинного перевода текста с одного языка на другой очень интересна, так как даже иногда у людей получается не познать основы языка верно, продолжая обучение на протяжении многих лет. К данной задаче есть примитивный подход - "скормить" нейронной сети большую выборку подготовленных данных. Однако есть теоретическая база для того, чтобы даже без огромного объема данных создать нейросетевую модель, которая сможет переводить фразы из одного языка в другой и наоборот благодаря идее о рекурсивных нейронных сетях. Основой исследования будем использовать метод seq2seq. Упрощенно, мы будем переводить текст в векторный вид, а потом учить модель сопоставлять в векторном пространстве два текста. В этой работе мы попытаемся создать такую модель для перевода с пары языков: русский - украинский.

2 Theory